Removal of Noise From Noise-Degraded Speech Signals

Panel on Removal of Noise From a Speech/Noise Signal

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704 0188

1a REPORT SECURITY CLASSIFICATION: Unclassified	1b RESTRICTIVE MARKINGS: N. A.
2a SECURITY CLASSIFICATION AUTHORITY: N. A.	3 DISTRIBUTION AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b DECLASSIFICATION DOWNGRADING SCHEDULE: N. A.	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: NAS-National Research Council; 6b OFFICE SYMBOL (If applicable): CHABA	7a NAME OF MONITORING ORGANIZATION: Office of Naval Research
6c ADDRESS (City, State, and ZIP Code): 2101 Constitution Ave., N. W. Washington, DC 20418	7b ADDRESS (City, State and ZIP Code): 800 N. Quincy Street Arlington, VA 22217-5000
8a NAME OF FUNDING SPONSORING ORGANIZATION: Office of Naval Research; 8b OFFICE SYMBOL (If applicable): Code 1142PS	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N0001487C0342
8c ADDRESS (City, State, and ZIP Code): Arlington, VA 22217-5000	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N 42	RR04209	RR0420901	4426124

11 TITLE (Include Security Classification)
(U) Removal of noise from noise-degraded signals

12 PERSONAL AUTHOR(S)
CHABA

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM 88/10/01 TO 89/09/30	89/06/01	34

16 SUPPLEMENTARY NOTATION

17 COSATI CODES		
FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Speech intelligibility; noise reduction, speech enhancement, speech testing, speech pereception.

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

(U) Techniques for the removal of noise from noise-degraded speech signals were reviewed and evaluated with special emphasis on live radio and telephone communications and the extraction of information from similar noisy recordings. The related area on the development of speech-enhancement devices for hearing-impaired people was reviewed. Evaluation techniques were reviewed to determine their suitability, particularly for the assessment of changes in the performance of workers who might use noise-reduction equipments on a daily basis in the applications cited above. The main conclusion was that noise-reduction methods may be useful in improving the performance of human operators who extract information from noisy speech meterial despite a lack of improvement found in using conventional closed-response intelligibility tests to assess those methods.

20 DISTRIBUTION AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL: John J. O'Hare	22b TELEPHONE (Include Area Code): (202) 696-4502; 22c OFFICE SYMBOL: Code 1142PS

Removal of Noise From Noise-Degraded Speech Signals

Panel on Removal of Noise From a Speech/Noise Signal

Committee on Hearing, Bioacoustics, and Biomechanics
Commission on Behavioral and Social Sciences and Education
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1989

This project was sponsored by the National Security Agency under contract number MDA904087-H-2009. The United States government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation hereon.

PANEL ON REMOVAL OF NOISE FROM A SPEECH/NOISE SIGNAL

JOHN MAKHOUL (Chair), Speech Signal Processing Department, BBN Laboratories, Inc., Cambridge, Mass.

THOMAS H. CRYSTAL, Division of Communications Research, Institute for Defense Analyses, Princeton, N.J.

DAVID M. GREEN, Department of Psychology, University of Florida

DOUGLAS HOGAN (deceased), Sparta, McLean, Va.

ROBERT J. McAULAY, Speech Systems Technology Division, Lincoln Laboratory, Massachusetts Institute of Technology

DAVID B. PISONI, Department of Psychology, Indiana University

ROBERT D. SORKIN, Department of Psychology, University of Florida

THOMAS G. STOCKHAM, Jr., Department of Electrical Engineering, University of Utah

Acknowledgments

The participants in the workshop on noise removal, conducted by the panel, gave presentations and participated in discussions that were very informative and that aided members materially in preparing this report. Workshop participants included Ernest Aschkenasy, Robert C. Bilger, Steven F. Boll, John Bunce, Douglas M. Chabries, Edward J. Cupples, Yariv Ephraim, Reger Freeyman, Jr., Daniel Graupe, Harry Levitt, Thomas J. Moore, Arlene Neuman, John Riney, Astrid Schmidt-Neilsen, and William Voiers.

The panel acknowledges the work of panel member Douglas Hogan, now deceased, who contributed significantly during the early stages of activity. We would also like to thank Edward P. Neuberg of the National Security Agency, who was very helpful to the panel in determining the objectives and scope of this effort.

Contents

Summary

The National Research Council's Committee on Hearing, Bioacoustics, and Biomechanics formed a panel to review and evaluate the effectiveness of techniques designed to remove noise from noise-degraded speech signals. This report describes both the techniques themselves and how they are currently evaluated. The panel surveyed the published literature and held a workshop of scientists and engineers active in the area. The panel was particularly concerned with applications to live radio or telephone communications and to the extraction of information from similarly noisy recordings, but it also reviewed the related area of developing and testing speech-enhancement devices for hearing-impaired people.

A number of noise reduction techniques have been developed that appear to reduce the perception of noise in the processed speech signal. However, these techniques have received only minimal acceptance and use, in part because standardized tests have shown that the intelligibility of the processed speech does not improve. For example, no improvement in intelligibility of processed speech has been demonstrated by closed-response tests such as the diagnostic rhyme test. Accordingly, evaluation techniques, such as intelligibility testing, were reviewed to determine their suitability, particularly for assessing changes in the performance of workers who might use such noise reduction equipment on a daily basis in the applications described above. The main conclusion of the report is that noise reduction methods may be useful in improving the performance of human operators who extract information from noisy speech material despite a lack of improvement found using conventional closed-response intelligibility tests. Such tests may not be appropriate for measuring the effectiveness of noise reduction systems. The report points to the importance of exploring the use of new ways to evaluate noise reduction methods for a variety of noise types and speech environments. For developing improved noise reduction methods, the report stresses the need for the appropriate use of short-term properties of speech signals and noise and for the development of perceptually derived design criteria that are based on the discovery and mathematical formulation of properties of human speech perception in noise.

Introduction

For over two decades, researchers have been investigating techniques for improving the quality and intelligibility of speech received in the presence of noise (Beek, Neuberg, and Hodge, 1977; Lim, 1983). Terms such as *noise removal*, *noise stripping*, *noise reduction*, and *speech enhancement* have been used to refer to such techniques. In this report, we shall use the term *noise reduction* to refer to techniques whose purpose is to reduce the perception of noise in a noisy speech signal. Noise reduction procedures received considerable attention about ten years ago, when digital computers made this type of speech processing practical. The refinement of these basic signal processing techniques is continuing. The advent of very-large-scale integration (VLSI) technology has also made it possible to incorporate noise reduction techniques in hearing aids.

Listeners generally agree that noisy speech processed by noise reduction techniques sounds less noisy and is easier to listen to (Lim, 1983). The sale of commercial noise reduction products in the last several years attests to the belief by users that such products are useful for their particular applications. Despite this, it would be fair to say that current noise reduction techniques have not been widely accepted for general application in the areas of speech communication, monitoring, and transcription. While there may be market-related reasons for the limited use of noise reduction devices (such as price), one important reason for the apparent lack of confidence in noise reduction techniques has been the fact that standardized intelligibility tests have consistently failed to show improvements in speech intelligibility when those techniques are used. In some cases a reduction in intelligibility has been reported (Lim, 1983). While informal reports of listener preference suggest that there may be performance benefits to be gained by installation of noise reduction systems, the negative quantitative findings of the more formal tests have made it difficult for many potential users to justify expenditures for system development and installation. The apparent contradiction between qualitative subjective impressions and quantitative findings from behavioral tests raises as many questions about the testing methods as about the speech processing techniques themselves.

To help resolve this apparent contradiction between listener impressions and formal test results, the Committee on Hearing, Bioacoustics, and Biomechanics (CHABA) of the National Research Council convened a panel of scientists and engineers who are expert in areas such as signal processing, speech in noise, psychoacoustics, experimental psychology, electronics, acoustical engineering, speech processing by computers, and telecommunications. The panel was charged with reviewing and evaluating the body of open literature to determine whether, in fact, intelligibility can be increased by processing the noisy speech signal and to indicate those modifications that appear to be most promising.

As part of its study, the panel held a workshop February 26-27, 1987, in Washington, D.C., to which a number of scientists and engineers expert in relevant areas were invited. The workshop and the panel's deliberations focused primarily on the relevant open literature, its validity, ar . its promise for future improvement of intelligibility of speech in noise. While anecdotal evidence was helpful to the panel, the emphasis was chiefly on information available in the open literature and on the formal testing of noise reduction techniques.

Although there are methods available that can help remove the effects of noise through the use of additional sensors and through the proper treatment of the acoustic environment of the transmission or recording, this report is concerned solely with single-microphone transmissions and recordings for which the acoustic environment is not under the listener's control. The topic of noise reduction under these conditions was reviewed previously by Beek et al. (1977) in the context of military communications and surveillance of communication channels.

Our review of the state of the art of removing noise from speech received in noise includes a critical examination of testing and evaluation procedures, as well as consideration of the structure and mathematical bases of the signal processing techniques. The review was conducted with the objective of identifying and discussing the issues related to actual jobs and tasks that involve listening to, and obtaining information from, a noisy speech signal.

The report is organized as follows. The section following this introduction gives a classification of application areas in which noise reduction can benefit the listener, followed by a section describing the different types of noise of interest. The next section reviews the various methods that have been developed for the reduction of noise in a noisy speech signal. The following section provides an overview of behavioral evaluation techniques for use in assessing the performance of noise reduction methods and includes a summary of reported test results for various noise reduction methods. The final section presents the conclusions of the panel on the current status of noise reduction techniques and their evaluation, followed by recommendations for future work in the development of new noise reduction methods as well as the development of evaluation techniques that can be used to assess the performance of noise reduction systems in the various application areas.

Classification of Application Areas

The various environments of interest and the measurement of human performance in those environments requires an understanding of the areas of application of noise reduction. Three such areas were considered by the panel:

(1) Two-way communication by voice,
(2) Transcription of a single, important recording, and
(3) Transcription of quantities of recorded material.

Of these, the third was considered the most important to the panel's charge.

The first area of application—noisy, two-way communication—is typified by flight-control communication between an airplane and a ground station. The communication activity consists of sequences of short messages. Noise may be present because of standard atmospheric radio interference, other transmitters using the same channel, and cockpit or tower audio interference picked up by the microphone. Noise reduction appears to be of benefit in this application because it would reduce the necessary concentration for the listening task by, one assumes, reducing listener fatigue and/or enabling the listener to focus attention more on the tasks of scheduling, routing, and collision avoidance. This application area may be characterized by a need for rapid response to spoken messages and by high task-related psychological stress. However, the listening part of the task is eased by the use of a highly redundant sublanguage with strict protocols or syntax and a limited vocabulary as well as by the two-way nature of the communication. The person receiving a message usually gives a confirmation and may, except in extremely urgent situations, ask for repeat transmission of questionable communications. Since in this application speech intelligibility is typically maintained by the language and protocol constraints, noise reduction would probably affect other aspects of job performance, such as listener fatigue. Testing of noise reduction for this application would need to take into account and possibly simulate all of the factors discussed.

The second area of application—transciption of a single, important recording—represents a very different application of noise reduction. By definition, there is one, important, noisy message to be understood. The analysis of a cockpit voice recorder following an airplane accident is one such application; the analysis of forensic material is another. The forensic recordings could come, for example, from the monitoring of the telephone lines of a law enforcement agency or from a hidden recorder. In all cases, the recording would be expected to contain a considerable amount of background audio noise and only a small amount of recorded material would need to be transcribed. The spoken material itself would be drawn from an unconstrained vocabulary and have a general grammar plus ungrammatical expressions. In analyzing this material, the transcriber is under little time pressure and may replay the recording many times and take repeated and long rests from the transcribing task. There is no task-related stress other than that of the concentration required for listening to noisy and distorted speech material. As related in the appendix to this report, written by one of the panel members who has extensive experience as an expert transcriber, an important ability is being able to listen to many different types of presentations of the same signal. These presentations usually involve different filterings and different playback speeds, but could also incorporate a noise reduction process. Some of the filtering for these presentations is analogous to part of the noise reduction process, but the latter employs complicated designs based on measurement of the actual noise. Hence, there probably are benefits to be obtained from using automated noise reduction techniques. Testing of noise reduction in this application would need to take into account the influence of the virtually unlimited opportunity for replaying and relistening to the original material.

The third area of application—transcription of quantities of recorded material—represents the most important area for this review of noise reduction. Two examples define possible sources for such materials and the tasks involved. In each, the material may be seen to be divided into messages or conversations of moderate length. Individual messages can differ in importance. The first example is that of a news and information agency that monitors and records the public broadcasts from a wide geographical area in a search for interesting, newsworthy developments. A second possible source of a similar volume of recorded material would be the monitoring of the telephone lines of a critical facility such as a nuclear power-generation station. The investigation following an incident could include making transcriptions of all conversations for the preceding weeks. Either radio interference or telephone line noise and background acoustic noise could be present in the recordings.

The characteristics of the task defined by these examples may be examined and compared with those of the tasks in the other two areas. Like the two-way communication task, transcription of quantities of recorded material is a full-time job, in which fatigue could seriously affect performance. However, it is a job in which there is no task-related stress other than that generated by the transcription process. There may be time pressure arising from the need to transcribe a certain quota of material. The requirement for accurate transcription can be addressed by replaying the recordings, but this must be balanced against the need to complete all the transcriptions in a reasonable time. The replaying process usually does not involve different presentations, such as different filterings, but could involve changes of speed. The spoken material has a large vocabulary, and both grammatical and ungrammatical expressions are common. The effects of noise reduction on this application should be evident by measuring the accuracy and volume of the transcriptions produced from test material of known content and specified noise conditions. Presumably, it would then be possible to develop tests for predicting the performance and impact of noise reduction procedures on actual transcription material and to devise procedures for validating the results of such predictive testing.

Noise reduction is also important to the users of hearing aids. The application has elements in common with the first and third application areas mentioned above. While the noise reduction techniques reviewed in this report might be used in hearing aids (in fact, certain noise reduction techniques have been developed especially for use in hearing aids), this application, with all its special considerations, is beyond the scope of this report.

Types of Noise

Definition of the signal processing task also requires specification of the types of noise to be reduced by processing. Not only do the different types of noise affect listeners differently, but also the noise reduction techniques depend on the types of noise. The description of noise types is limited to those that commonly occur and that have been addressed, with some qualitative success, by noise reduction techniques. It should be noted that the techniques themselves modify the speech as well as the noise. Noise reduction is therefore often achieved at the cost of some speech distortion or the introduction of a different, lower-amplitude noise.

The major types of noise that have been addressed by noise reduction methods can be classified as either impulsive or continuous. Continuous noise can be further subclassified as either wideband or narrowband (including tones).

Impulse noise is characterized by the occurrence of additions to the signal that have durations not exceeding several tenths of milliseconds (ms) and that are separated by longer time intervals. Impulse noise is further characterized as nonrhythmic, occurring at unpredictable times, and having a variable signal shape. (Although it is possible that there is impulse noise that occurs at fixed, predictable intervals or having fixed, measurable signal shapes, it has not been identified as being significant in speech processing.) Impulse noise can occur as a result of sudden, brief atmospheric disturbances or switching of the characteristics of transmission equipment. It is typically heard as clicks superimposed on the speech, even at very low amplitudes. However, noise reduction techniques have addressed only the removal of impulse noise whose amplitude is larger than the neighboring speech signal.

Continuous noise is characterized by the fact that it is present on a continuing basis and that its characteristics change slowly relative to variations in the speech signal. Different types of continuous noise are usually described in terms of their frequency or spectral characteristics. Narrowband continuous noise typically occurs as tones whose frequencies and amplitudes are slowly varying relative to the variation of the spectrum of the speech signal. Tonal inference can arise from, among other things, competing transmitters in a broadcast channel, malfunctions of radio or telephone transmission equipment, excessive audio feedback, or machinery that generates acoustic noise.

Broadband continuous noise is characterized by having energy in a large band of frequencies that effectively overlaps a major part of the speech spectrum. This type of noise is typically random in nature and can arise electrically from thermionic sources in the atmosphere or in the equipment being used, from the combined occurrence of electrical disturbances, such as lightning strikes, or the combined effect of competing equipment in the same communications channel. Noises that are individually impulsive, tonal, or otherwise structured can appear in the aggregate to be continuous and unstructured. This type of noise can also arise acoustically from wind near the microphone, irregular operation of machinery, or summed structured acoustic sources. Background voices or babble, which may be considered to be a noise of this type, is a particularly significant noise for hearing aid development.

Because broadband continuous noise overlaps the speech signal in both time and frequency, it is the most difficult type of noise to remove. Unfortunately, it is also the most ubiquitous type of noise and has received the greatest attention in the development of noise reduction techniques.

Methods for Noise Reduction

In the past two decades, a number of different methods have been developed with the aim of reducing the perceived noisiness and increasing the intelligibility of noise-corrupted speech signals. Many of these methods are contained in a set of papers collected by Lim (1983). The tutorial by Lim and Oppenheim (1979) presents a good review of the literature on noise reduction methods developed prior to 1979.

In this section we present a brief overview of the major techniques employed in the various noise reduction methods, with emphasis on methods that have found their way to real-time implementation. Since most of the methods have been developed to deal chiefly with slowly varying continuous noise, broadband as well as narrowband, much of the discussion focuses on that topic. Methods for reducing continuous noise can be classified generally into frequency-domain and time-domain approaches, and they are discussed in separate sections below. In contrast, the reduction of impulse noise has received relatively little attention; the relevant literature in that area is reviewed in a third section below.

FREQUENCY-DOMAIN METHODS

Overall Approach

Figure 1 shows a canonical diagram for frequency-domain noise reduction. The incoming noisy speech signal is analyzed on a short-term basis, one block at a time. A block, which is typically in the range of 20 to 40 ms, is known as a frame. Each frame of noisy speech is spectrally decomposed (either by a discrete Fourier transform or a bank of bandpass filters), into a set of magnitudes and a set of associated phases, each magnitude-phase pair corresponding to a distinct frequency component. The noise reduction is achieved by appropriate adjustment of the set of spectral magnitudes. A waveform reconstruction process then combines the adjusted magnitudes with the nonmodified phases (through either an inverse Fourier transform or a set of bandpass filters), resulting in a signal that can be viewed as an estimate of the noncorrupted speech signal. The whole process is repeated for each frame of input signal.

The various noise reduction methods differ primarily in their approaches to spectral magnitude adjustment. All methods typically employ a speech activity detector, which detects when speech is present. If a decision that speech is not present is made, then one can estimate the spectrum of the background noise. This estimate is updated on a continuing basis to reflect possible changes in the background noise. For each frame of input

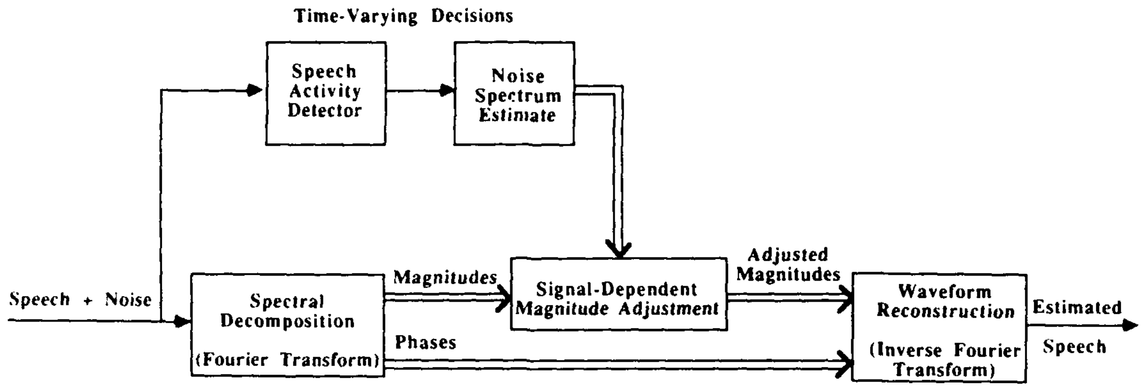

FIGURE 1 Canonical diagram for frequency-domain noise reduction techniques. (Double lines indicate a set of component values.)

signal, the then-current estimate of the noise spectrum is employed in some appropriate adjustment of the signal spectrum.

The different methods for signal spectral magnitude adjustment can be divided into two types: frequency-selective methods and transform-domain methods. Frequency-selective methods adjust the spectral magnitudes on a frequency-specific basis, i.e., the adjustment is performed at each frequency separately, while transform-domain methods perform the adjustment indirectly in a domain that is a mathematical transform of the spectrum. The two types of methods are described further below.

Frequency-Selective Methods

In frequency-selective methods, a speech-to-noise power ratio is computed at each frequency component, using the current estimate of the noise spectrum and the speech spectrum for the input frame. The signal spectral magnitude is then attenuated for each frequency component by an amount determined by a precomputed characteristic that gives the amount of attenuation for each speech-to-noise ratio. As a result, different frequency components are generally attenuated by different amounts, depending on the speech-to-noise ratio computed for each frequency component.

Because of their frequency-selective property, it is possible to use these methods to perform noise reduction with broadband noise as well as narrowband noise, including time-varying tones.

A number of approaches have been taken to derive suitable relations between magnitude attenuation and speech-to-noise ratio. In particular, the method of spectral power subtraction has been developed by several researchers (see, for example, Schroeder and Noll, 1965; Lim, 1978; Berouti, Schwartz, and Makhoul, 1978; Boll, 1979; Preuss, 1979). McAulay and Malpass (1980) have developed techniques based on Wiener filtering and maximum likelihood estimation; improvements in the latter approach have been made by Ephraim and Malah (1984). Examples of some of these attenuation versus speech-to-noise ratio plots are shown in Figures 2 and 3. In general, as the speech-to-noise ratio decreases, i.e., the power of the speech decreases relative to the power of the noise at each frequency of interest, the attenuation of the spectral magnitude at that frequency is increased. In this fashion, regions in which the noise power is relatively more dominant are attenuated more than regions in which the speech power is more dominant. The result of this process is a reduction in the perceived level of noise.

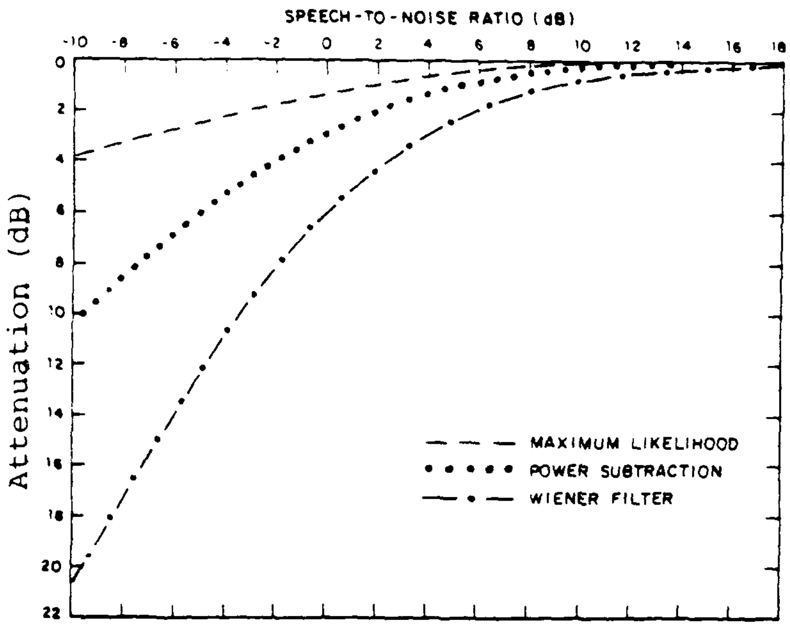

FIGURE 2 Maximum likelihood, power subtraction, and Wiener filter plots of spectral attenuation versus frequency-specific speech-to-noise ratio. Source: McAulay and Malpass (1980). Copyright © IEEE. Reprinted by permission.

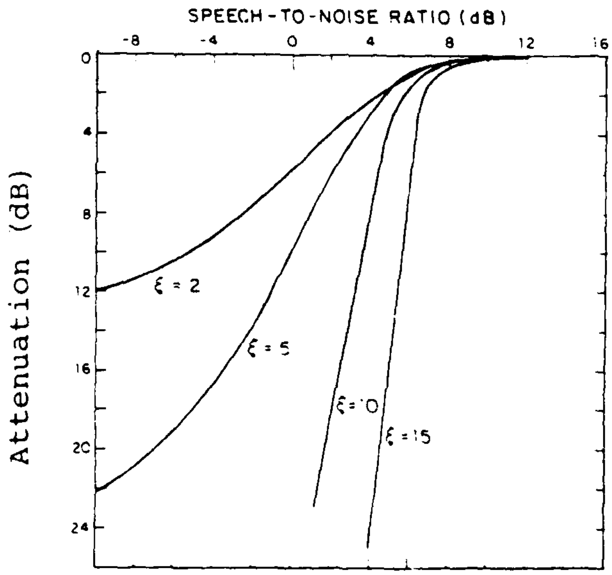

FIGURE 3 Parametric attenuation plots for the soft-decision maximum likelihood approach. Source: McAulay and Malpass (1980). Copyright © IEEE. Reprinted by permission.

In addition to significant reduction in the perception of noise, the various approaches often introduce various types of low-level distortions in the signal. Most common is a low-level noise that has a musical or tonal quality that can be annoying to the listener. Several researchers have attempted to maintain high speech quality by using less attenuation while adding low-level white noise to the output to mask the musicality of the processed noise (Berouti et al., 1979; Ephraim and Malah, 1984).

Heavy attenuation of spectral magnitudes can completely eliminate the perception of noise—but at the cost of reduced speech intelligibility, principally through the severe attenuation of low-energy speech sounds, such as consonants. Typically, some compromise is made between noise reduction and possible loss in speech quality or intelligibility. The issue of measuring possible enhancement in speech intelligibility is addressed below in the section on assessing human performance.

The effectiveness of noise reduction methods can be enhanced by exploiting certain time-dependent properties of the speech and noise. In addition to speech activity detection, which depends on the assumption that the noise statistics change more slowly in time than do those of the speech, improved speech quality can be effected by smoothing the spectral magnitudes in time in a manner that depends on whether the speech power level is rising or falling. An increasing speech level may reflect the onset of a speech event; hence, the smoothing time constant is shortened to capture better any low-energy speech information, such as in an initial consonant. A decreasing speech level uses a longer time constant to prevent any low-level trailing speech energy from being attenuated.

Transform-Domain Methods

In transform-domain methods, the signal spectrum is first transformed to another domain in which the noise reduction processing takes place. The most notable example of these methods is the one used in the INTEL system developed by Weiss and Aschkenasy (1975), which began with an effort that predated the development of frequency-selective methods. In the INTEL method, the signal spectral magnitudes are first compressed by taking the fourth root of each magnitude (Aschkenasy, 1986). (Various types of compression were attempted, including logarithmic compression and various nth-root compression rules, where n was varied over a wide range, but fourth root compresion yielded the best compromise between noise reduction with minimal speech distortion and real-time processing.) The compressed magnitudes are then transformed via a Fourier transform into a set of ordered transform coefficients. (For logarithmic compression, these coefficients are known as cepstral coefficients [Oppenheim and Schafer, 1975] but no standard name exists for arbitary nth-root compression.) Because broadband noise is typically concentrated in the low-order transform coefficients while speech is spread out over a wide range of coefficients, noise reduction is effected by attenuating the low-order transform coefficients through a transform subtraction process, whereby a weighted estimate of the noise transform is subtracted from the signal transform, taking care to retain the sign of each coefficient. (The weighting is higher for low-order coefficients by a factor of two approximately.) An inverse Fourier transform followed by a process of expansion (by taking the fourth power for fourth-root compression) yields the set of adjusted spectral magnitudes, which are then combined with the original phases to obtain the output processed speech (see Figure 1). The process also includes a smoothing technique that restores much of the power contour of the signal.

The INTEL process has been effective in reducing the perception of broadband noise. However, the method was not intended and, in fact, cannot be used to reduce the perception of narrowband noise, including tones. The reason is that, unlike broadband noise, narrowband noise affects the whole range of transform coefficients, not just the low-order coefficients. A separate frequency-selective method is used, therefore, to deal with narrowband noise (Aschkenasy, 1986).

A different transform-domain method was developed by Suzuki and others (Suzuki, 1976; Suzuki, Igarashi, and Ishii, 1977; Nakatsui, 1979) whereby a Fourier transform of the signal power spectrum (the square of the spectral magnitudes) is taken first, resulting in an autocorrelation sequence. The low-order coefficients of this sequence are then attenuated to reduce the effects of broadband noise. Here the method diverges from that shown in Figure 1 in that the waveform reconstruction process is performed by splicing sections of the autocorrelation sequences of consecutive frames, followed by a spectral square-rooting process to maintain the proper spectral magnitudes. In effect, this method also modifies the short-term phase. The question of the possible benefits of phase modification is discussed next.

The Role of Phase

The canonical method for noise reduction shown in Figure 1 depends on adjusting the magnitudes of the short-term signal spectrum while keeping the phases the same. The question arises as to whether some adjustment of the short-term phase could aid in the noise reduction process. Theoretical arguments based on simplifying assumptions that model speech as a Gaussian random process show that keeping the short-term phase the same is the best one can do in estimating the speech from a noisy speech signal with additive Gaussian noise (Ephraim and Malah, 1984).

In addition to the theoretical arguments that justified the focus on adjusting the magnitudes and not the phases of the short-term spectra, Weiss and Aschkenasy (1975), Wang and Lim (1982), and Ephraim and Malah (1984) performed controlled experiments to determine the relative perceptual importance of modifying the magnitudes and phases of the short-term spectrum. In one experiment, the spectral magnitudes of the noisy speech were not modified at all but the phases were set equal to those of the corresponding clean speech; the result of the waveform reconstruction was a signal that was perceptually the same as the noisy speech, with no perception of noise reduction. In the second experiment, the reverse was done: the spectral magnitudes were set equal to those of the corresponding clean speech, but the phases were kept unmodified as in Figure 1. The result in this case was a signal that was perceptually similar to the clean speech, with complete elimination of the noise.

Theoretical as well as experimental evidence, therefore, points to spectral magnitude adjustment as the primary vehicle for noise reduction, with phase adjustment having little or no effect on noise reduction.

It should not be concluded from the above that the short-term phase is unimportant and therefore could be set to arbitrary values. Indeed, one could adjust the phase to distort the signal drastically, resulting in reduced quality and intelligibility. The correct conclusion is that by adjusting only the phase, one cannot hope to improve the quality of the noisy speech signal in any substantial way.

TIME-DOMAIN METHODS

The canonical *frequency-domain noise reduction* process depicted in Figure 1 can be viewed equivalently as a time-varying linear filtering process, wherein the spectral characteristics of the linear filter for each frame of input signal depends on the spectrum of the signal as well as the estimated spectrum of the noise. Since this linear filter does not adjust the phases of the input signal, it can be considered to have linear phase (or constant delay) for all frequency values and for all time. In several of the frequency-selective methods

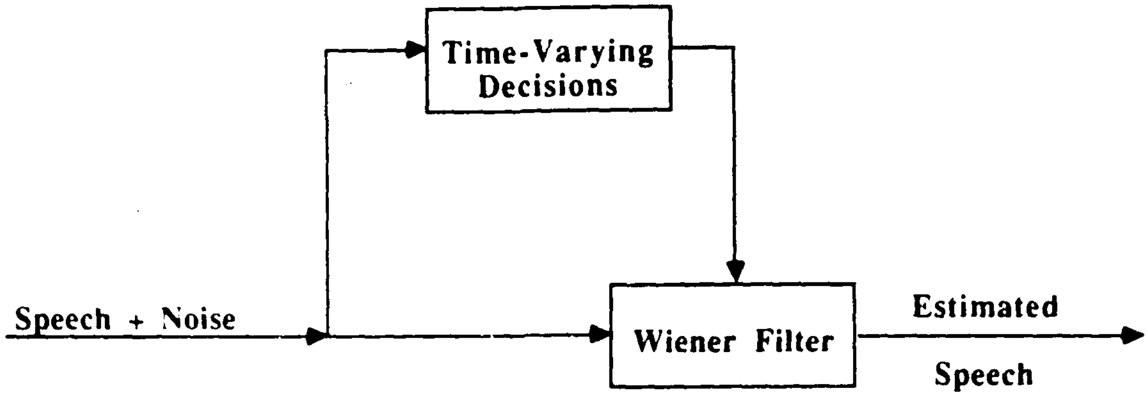

FIGURE 4 A time-domain canonical implementation of a minimum mean-squared error Wiener filter for noise reduction.

mentioned above, the linear filter is computed as the one that attempts to minimize the minimum mean-squared error between the estimated speech and the clean speech signal. Such a filter, known as a Wiener filter (Van Trees, 1968), can be implemented in the frequency domain as in Figure 1. A time-domain canonical implementation of the same filter is shown in Figure 4. In this case, the Wiener filter operates on the input signal on a sample-by-sample basis. As in Figure 1, the system in Figure 4 employs time-varying decisions, including speech activity detection and the estimation of the noise spectrum.

This time-domain approach has been taken by Graupe et al. (1987), who has developed noise reduction techniques for hearing aid applications. Separate parameter vectors for modeling the speech and the noise are estimated and used to design the Wiener filter. The Wiener filter is used whenever a decision is made that noise is present; if a no-noise decision is made, then the output is set equal to the input. Graupe states that the filter design further employs heuristics based on the characteristics of speech sounds. Although the general theory of Graupe's approach has been presented (Graupe, 1984), a great deal of the detailed information needed to implement his techniques has not been published.

One approximation to the Wiener filter that has been used in noise reduction applications is Widrow's least-mean-squares (LMS) method (Widrow et al., 1975). Usually, this method requires a second microphone to receive a correlated measurement of the noise process, but, owing to the fact that speech is highly correlated in time, the second input can be approximated by a delayed version of the noisy input, as shown in Figure 5. A commercial product based on this approach has been developed by Paul (1978). In some implementations of the LMS approach (Sambur, 1978; Veenemand and Mazor, 1987), the delay is taken to be a pitch period, but this approach can lead to speech distortion during pitch changes and, more significantly, can result in the attenuation of unvoiced speech (as in consonants such as f, s, p, t, k). Furthermore, this approach requires knowledge of the pitch, which in itself is difficult to estimate in noise. Chabries et al. (1982, 1987) have tried to reduce these effects by using a delay of less than 0.5 ms, a duration over which the speech samples should be highly correlated. However, effective noise cancellation then depends on the assumption that the noise is uncorrelated over this short duration, which decreases the usefulness of the approach if the noise is narrowband or if slowly varying tones are present.

Even though the inner workings of time-domain methods are generally not as transparent as frequency-domain methods in their effect on the signal spectrum, the fundamental

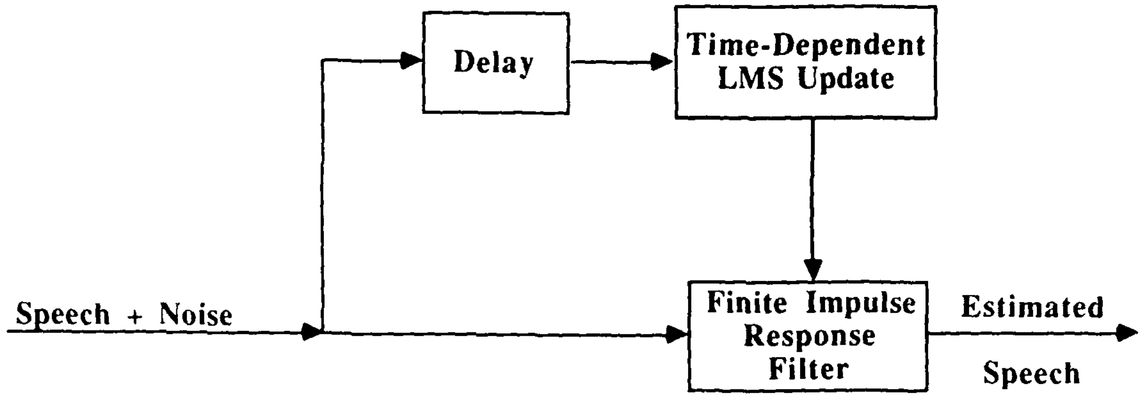

FIGURE 5 A canonical single-input time-domain least-mean-squares (LMS) filter for noise reduction.

nature of what they are trying to accomplish is very similar. One salient difference between the two sets of methods, however, is the treatment of phase. In contrast with the frequency-domain methods mentioned above in which the short-term signal phase is not modified, time-domain methods typically modify the phase as well as the signal spectrum. However, whatever phase modification takes place is more a by-product of the time-domain operations than a deliberate attempt to manipulate the phase in the hope of reducing the noise. The remarks made above concerning the role of phase in noise reduction should also apply to time-domain methods.

IMPULSE NOISE REDUCTION

Previous sections described methods that were designed to reduce the effects of broadband noise and narrowband noise, including slowly varying tones. These methods are generally ineffectual against impulsive or intermittent noise. Therefore, a different approach is needed to reduce the perception of this type of noise. A number of methods have been developed in the context of reducing the effects of channel errors in digital transmission of speech, which is similar to the problem of removing impulse noise (Steele and Goodman, 1977; Kundu and Mitra, 1987). Weiss and Aschkenasy (1978) have used similar techniques to remove impulsive interference occurring in a speech signal.

Ideally, it is desired to filter only those points in the input signal that are corrupted by impulse noise and leave the uncorrupted data points as they are. In general, this process requires a technique for thresholding and, if the threshold is passed, the data point is declared to be corrupted by noise and is subsequently filtered. The filtering usually involves a deletion of the noisy data and an interpolation to replace the deleted region by some speechlike waveform. A block diagram for a canonical impulse noise reduction procedure is shown in Figure 6.

CONCLUSIONS

Different noise reduction methods have been designed to be most effective for different types of interference. Reduction of impulsive noise, for example, requires the use of time-selective modification of signal values, while slowly varying continuous noise (broadband and narrowband) is obtained chiefly through adjustment of the magnitude of the short-term signal spectrum. Modification of the short-term phase does not appear to be especially useful in the reduction of continuous noise.

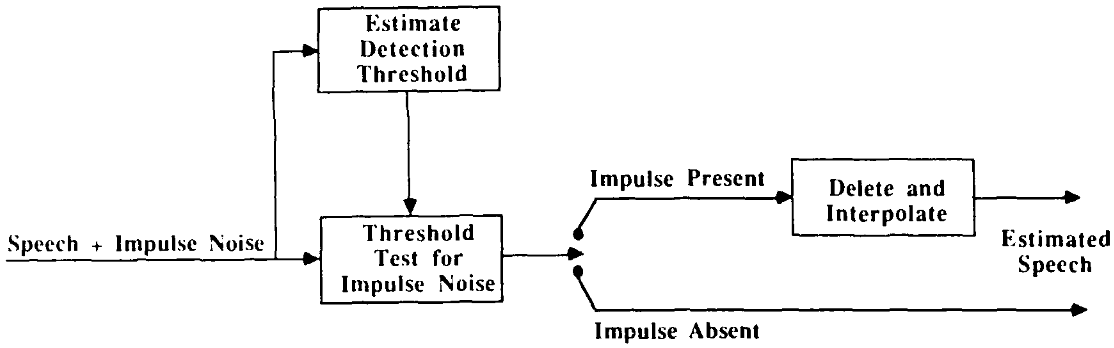

FIGURE 6 A canonical implementation for reduction of impulse noise.

Generally, mathematical and statistical criteria have been used in the design of noise reduction methods with little regard for perceptual criteria derived from listening tests with human subjects. However, additional time-varying decisions that depend on knowledge of properties of speech and noise have been fruitfully incorporated. Examples include speech activity detection, estimation of the noise spectrum, and voiced/unvoiced decisions.

It is not surprising that the effectiveness of any noise reduction technique can be enhanced through the judicious application of properties of speech signals and noise and of human auditory perception. Although many noise reduction techniques have been based on one or another mathematical model of speech and noise, it is through research aimed at exploiting additional properties of speech and auditory perception that future improvements in the effectiveness of noise reduction techniques will be expected to emerge. The needed perceptual knowledge will require additional research effort into understanding further the perception of speech in noise by human listeners, as well as the formulation of such knowledge in appropriate mathematical form for use in the design of perceptually based noise reduction techniques.

Assessing Human Performance

The effectiveness of noise reduction techniques depends ultimately on their utility for human listeners. Upon examination of the existing literature on noise reduction methods, it is apparent that, although a good deal of engineering effort has gone into the design and implementation of noise reduction methods, there has been a conspicuous absence of parallel efforts aimed at the formal and controlled evaluation of these methods and devices using human listeners. Quantitative data on changes in speech intelligibility, comprehension, and operator workload when a particulr noise reduction method is used are often not available in the published literature. The limited test data that are currently available are unfortunately equivocal. Although the data show that speech quality may be improved by noise reduction techniques, parallel improvements in speech intelligibility have not been observed in any of the formal listening tests reviewed by the panel. In some cases, the data show decreases in performance with a particular algorithm. There is the suggestion that the failure to find improvements in speech intelligibility may be more a result of the particular tests used to measure speech intelligibility than a limitation of the noise reduction methods tested (Schmidt-Nielsen, 1987). Developing more discriminating performance measures and retesting promising noise reduction techniques using those measures may assist in advancing the state of the art in this area.

In an attempt to understand better the results obtained in testing noise reduction techniques and to aid in the formulation of appropriate recommendations for future work, the panel felt it was important to consider first several of the major factors that have been shown to influence human performance in speech communication tasks. A summary of these factors is given in the next section. Following that is a review of the most common behavioral measures or tests currently used to assess speech intelligibility, quality, comprehension, communicability, and listener fatigue and workload. Then we summarize published results for several noise reduction methods that were tested in a small number of formal and informal studies.

The human listening tests used in evaluating noise reduction methods have been limited in scope. Although many of the systems developed have initially reported improvements in intelligibility, more rigorous testing under controlled laboratory conditions has generally yielded negative findings. Despite the limited and generally negative nature of these results, we end this discussion with some general conclusions about the effectiveness of these methods and about the appropriateness of existing perceptual tests for measuring speech intelligibility in noise. In some cases, our conclusions are based in part on unpublished experimental data, on anecdotal evidence, and on the testimony of highly experienced listeners who have used a variety of techniques to help identify degraded speech signals.

FACTORS AFFECTING HUMAN PERFORMANCE

The performance of human listeners in any speech communication task is affected by a number of perceptual and cognitive factors (Pisoni, Nusbaum, and Greene, 1985). To provide a framework for interpreting the results of human performance studies, we first consider a number of variables, excluding speech-to-noise ratio, that may affect an observer's performance in speech communication tasks: (1) the specific requirements of the task, (2) the physiological and anatomical limitations of the human observer, (3) the experience and training of the human listener, (4) the message characteristics, (5) the structure of the speech signal, and (6) secondary or indirect effects of noise on performance. These factors often overlap and they interact in various ways to affect human performance.

Task Requirements

In some tasks the demands are relatively simple, such as deciding which of two known words was spoken. Other tasks are extremely complex, such as trying to recognize an unknown utterance from a virtually unlimited number of response alternatives. In addition to the primary communication task, the listener may also be engaged in some activities that already require substantial effort and processing resources. There is research in the cognitive psychology and human factors literature demonstrating the powerful effects of perceptual set, instructions, subjective expectancies, cognitive load, and response set on performance in a variety of perceptual and cognitive tasks (Wickens, 1984). The amount of context and the degree of uncertainty in the task also strongly affect an observer's performance in substantial ways (Kantowitz and Sorkin, 1983). Thus, it is necessary to understand the requirements and demands of a particular task before drawing any strong inferences about the listener's performance or about the utility of a particular speech processing technique. At the present time, the panel was unable to find any studies that investigate the interactions between these task variables and the effects of different noise reduction algorithms.

Physiological and Anatomical Limitations of the Observer

The second factor influencing perception of speech concerns the physiological and anatomical limitations on the human's ability to input, encode, store, and retrieve information. Because the nervous system cannot maintain all aspects of sensory stimulation, there are constraints on the human observer's capacity to input and encode raw sensory data. The listener must rapidly transform the input sensory data into abstract internal codes suitable for temporary storage in the working or short-term memory stem (Baddeley, 1986). Encoded data from different sensory channels must pass through this limited capacity system before further information processing. Finally, information that has been sufficiently processed in short-term memory may be transferred to the long-term memory system for later retrieval. The bulk of research in perception and cognitive processes over the past 25 years has identified the short-term memory system as the major bottleneck in the internal flow of information (Shiffrin, 1976). The amount of informatio that can be processed and held in short-term memory is severely limited by the listener's attentional state, past experience, and the quality of the original sensory input. These processing limitations interact with other factors to affect the listener's performance in a particular task.

Experience and Training

The third factor concerns the ability of human observers to learn rapidly effective cognitive and perceptual strategies to improve their performance in almost any task. When given appropriate feedback and training, subjects can learn to classify novel stimuli, remember complex stimulus sequences, and respond to rapidly changing stimulus patterns in different sensory modalities (Watson, Kelly, and Wroton, 1976; Kidd, Mason, and Green 1986; Nickerson and Freeman, 1974). Clearly, the flexibility of subjects in adapting to the specific demands of a task is an important factor that must be considered in attempts to evaluate the effectiveness of any speech processing technique. Some changes may be dramatic, whereas others may require a more gradual period of perceptual learning and adaptation. One member of the panel suggested that some processing techniques may produce improvements in the performance of relatively inexperienced listeners, while other techniques may be of help only to highly experienced listeners who have had many hours of exposure to degraded speech signals. Apparently this issue has not been addressed in the published literature, although it is obviously relevant to the current problem.

Message Characteristics

The fourth factor concerns the constraints on the number of possible messages and the organization and linguistic properties of the messages. (This is clearly a major component of the general task requirements factor and is one of the better-studied areas of speech communication.) We summarize this factor by referring to the totality of all potential messages as the message set. A message set may consist of words that are distinguished only by a single phoneme, or it may consist of words and phrases of very different lengths, stress patterns, and phonetic variability. Use of these constraints by listeners depends on prior linguistic knowledge (Miller, Heise, and Lichten, 1951). The choice and arrangement of speech sounds into words is constrained by rules of allowable sound sequences in a language; the arrangement of words in sentences is constrained by the grammar of the language; and finally, the meaning of individual words and the overall meaning of sentences in a text is constrained by the set of concepts that can be communicated in that language. The contribution of these various sources of knowledge to speech perception will vary substantially from isolated words, to sentences, to passages of fluent continuous speech. The effects on performance due to the characteristics of the message set are shown in Figure 7, which is a plot of intelligibility (in percentage correct) as a function of the speech-to-noise level for different message sets (Webster, 1978). At a given speech-to-noise level, increases in the predictability of the materials yield improved intelligibility. The top horizontal scale of the figure shows a transformation of the speech-to-noise ratio called the articulation index. The articulation index is a procedure used to predict the effects of noise on speech by considering the weighted speech-to-noise levels in a set of frequency bands spaced across the speech spectrum (French and Steinberg, 1949; American National Standards Institute, 1969).

Structure of the Speech Signal

The fifth factor refers to the physical structure of the speech signal itself. Speech signals may be thought of as the acoustic realization of a complex and hierarchically organized system of linguistic rules that map concepts into sounds. The acoustic properties of the speech signal are constrained in substantial ways by vocal tract acoustics and articulation (Stevens, 1964) and through the intonation of different linguistics structures (for example,

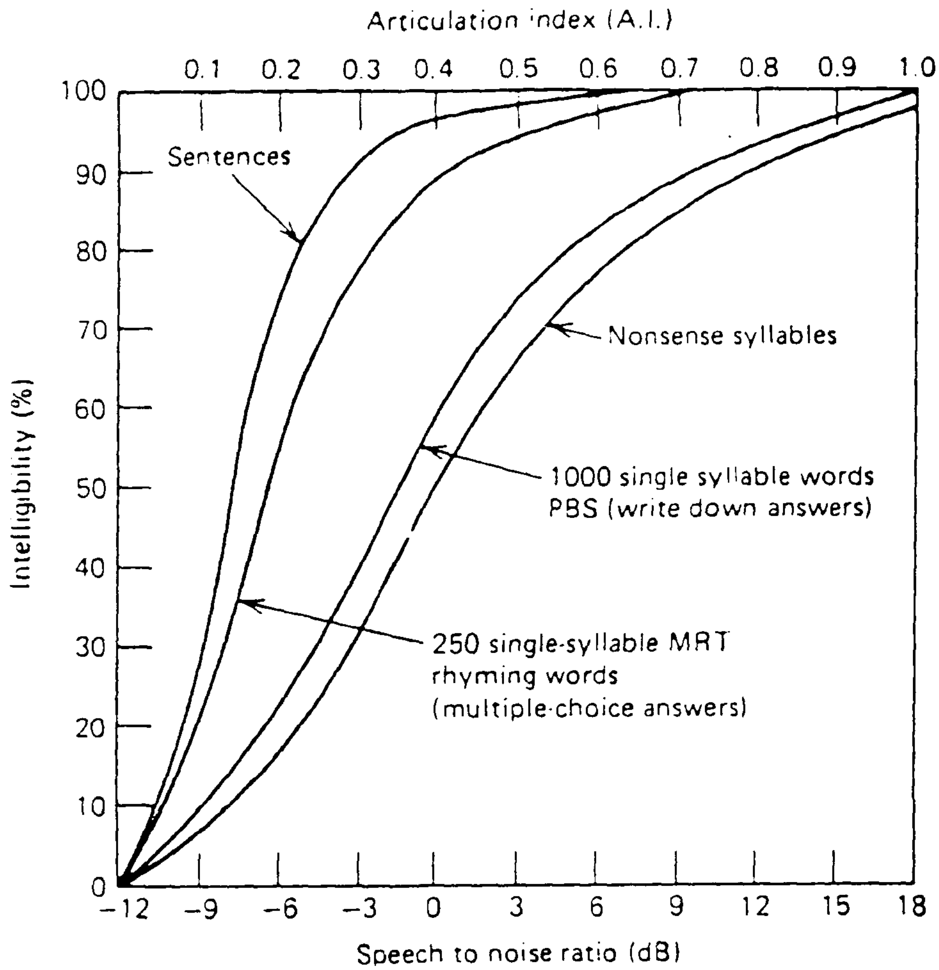

FIGURE 7 Intelligibility of different test materials as a function of speech-to-noise ratio. Source: Webster (1978). Reprinted by permission.

questions versus statements). Speech degraded by noise represents an impoverished acoustic signal that contains only a limited subset of the normal information-bearing elements of speech (Kryter, 1985).

Secondary or Indirect Effects

Apart from its effect on the intelligibility of speech, a noisy environment may produce secondary or indirect effects on the performance of some tasks. Removing noise from a noisy speech signal may result in an improvement in the comfort or performance of a listener who is required to work for an extended listening period. However, it is not known whether long-term performance improvement can occur in the absence of observed changes in traditional measures of intelligibility. The panel was unable to find any published research that addresses the issue of long-term improvement in human performance using noise reduction techniques. Investigators studying the effects of noise reduction on performance have suggested several hypotheses about the possible mechanisms involved in improving human performance (Hockey, 1983; Kryter, 1985; Loeb, 1986):

(a) Reduction of some interference with "inner speech" (interference with working memory or rehearsal processes) or other cognitive processes.

(b) Modification of the information-processing strategies of the worker, for example, by changing the worker's attentional or observational strategies, speed-accuracy trade-off, or response criteria.

(c) Reduction of fatigue, arousal, distraction, and other effects related to noise as a stressor.

These hypotheses are not unreasonable; the noise literature includes some studies that support and others that fail to suggest each hypothesis. Unfortunately, given the present base of published data, it is not possible to draw conclusions about the validity of any of the hypotheses with regard to the usefulness of noise reduction algorithms.

EVALUATION OF SPEECH COMMUNICATION SYSTEMS

A number of well-established techniques for the evaluation of speech communication systems are presently available. In reviewing these techniques and their applicability to the evaluation of noise reduction devices, it is important to recognize that these evaluation techniques differ on an objective-subjective dimension. The words *objective* and *subjective* have widely different meanings to people with different backgrounds. Many engineers, for example, classify any evaluation technique involving human listeners as subjective; the term *objective* is reserved for evaluation techniques that are performed by mechanical means, such as a measuring instrument or a computer. To behavioral scientists, both terms are used to describe evaluation techniques that employ human observers: objective methods are those for which the responses of the human observer can be scored as correct or incorrect against a standard, while subjective methods elicit a judgment of preference that cannot be scored as correct or incorrect. In this report, when discussing tests that involve human listeners, our use of the terms *subjective* and *objective* conforms to the usage by behavioral scientists.

At one extreme, a communication system can be evaluated by measuring some physical parameters of the system (bandwidth, speech-to-noise ratio) and an evaluation index can be calculated, based solely on these physical measurements. Two examples of this approach are the articulation index (AI), first developed by French and Steinberg (1949) and later refined by Kryter (1962), and the speech transmission index (STI) of Steeneken and Houtgast (1980). At the other extreme, one can simply ask the listeners to evaluate the quality of a communication system. Such opinions can be quantified to some degree by using a rating scale or by counting the number of times one system is preferred to another. The basic response, however, is purely subjective; it is simply a personal preference of the listener. Between these two extremes are a number of well-established objective evaluation methods used by behavioral scientists to evaluate the performance of speech communication systems. A prime example is an intelligibility test, in which a list of words is transmitted and the percentage of words correctly understood by a human observer is the primary measure of performance. Such an evaluation technique is clearly not simply a matter of opinion; it reflects how well the human listener can discriminate or identify a set of speech signals heard through the communication system under evaluation.

In part the problem one faces with many noise reduction devices is that there is a conflict between the purely subjective response and the intelligibility score results. Speech passed through the system sounds less noisy and may therefore be preferred in terms of a purely subjective response. The device does not, however, produce a higher score on the less subjective test in terms of the number of words correctly understood. Interpreting and understanding this conflict was one of the primary problems faced by the panel.

Of course, the valid evaluation of a system's effectiveness may be attainable only by studying actual users operating in the intended environment. Most of the experimental data with noise reduction systems has been obtained in laboratory situations that may not accurately emulate all the relevant operator, system, and environmental variables. To some extent, measures of system performance, whether or not based on objective measurements, are only as good as they can accurately predict performance in the real application environment.

We list below the types of techniques that have been employed in evaluating speech communication systems. They are discussed starting with the simpler procedures and proceeding to the more complex. The first three procedures are all scorable in terms of a correct or incorrect response; the last procedure is purely subjective.

Speech Intelligibility Tests

The primary goal of speech communication is for the listener to identify the message produced by the speaker. The speech intelligibility test measures how well this goal is achieved. The speaker reads from a set of messages and the listener responds by transcribing the message. The intelligibility test score is the percentage of correctly received messages. An important variable in determining this score, as discussed earlier, is the size and structure of the potential set of messages. The message set used in most evaluations may be composed of isolated syllables, words, or sentences.

A variable of considerable importance in determining performance is whether the listener knows the set of potential message alternatives. If the listener knows items in the response set, we say the message set is closed. In that case, the listener is forced to select from a limited set of response alternatives on each trial. The set of messages can be less restricted, for example, to include all English words. In that case, we say the message set is open. This closed/open distinction is of importance for the present problem, since almost all speech intelligibility tests used in the early evaluations of noise reduction systems employed closed message sets (usually consisting of only two response alternatives).

The most common intelligibility tests use isolated words and closed response sets. Examples of these tests include the modified rhyme test (MRT), which uses six alternatives that differ by a single consonant in either initial or final position (House et al., 1965) and the diagnostic rhyme test (DRT), which has two-alternative responses on each trial consisting of pairs of words that differ by one distinctive feature in the initial consonant (Voiers, 1977b, 1983). Because the answer sheets in these tests are multiple choice, the influence of the experience and training factor (discussed above) on the testing situation is minimized.

Other intelligibility tests present monosyllablic words in isolation or in standard carrier phrases using an open-response format in which the listener repeats the word heard on each trial. Examples of such open-response tests include word lists developed at the Central Institute for the Deaf (Hirsh et al., 1952) and at Northwestern University (Tillman and Carhart, 1966). In contrast to closed-response tests, open-response tests require the listener to have a greater natural language ability, but such abilities may not correspond to those needed to perform a specific task.

Intelligibility scores have also been obtained for words presented in sentences as well as phonemes in isolated consonant-vowel or vowel-consonant nonsense syllables. In the latter case, confusion matrices can be generated and used for diagnostic purposes in identifying specific problems in the communication system (Miller and Nicely, 1955; Wang and Bilger, 1973).

More recent efforts have been directed at the development of sets of sentences that differ in predictability, such as the speech perception in noise (SPIN) test (Kalikow, Stevens, and Elliott, 1977) or sentences in which there is a predominance of specific phonemes in various phonetic environments, such as the phoneme specific sentences test (Huggins and Nickerson, 1985). A property of all these intelligibility tests is that they involve measures of performance expressed as a percentage of correct responses. In all cases, subjects are required to transcribe or respond to the acoustic-phonetic properties of the speech signal (see also Picheny, Durlach, and Braida, 1985). It is assumed that these minimal properties contribute substantially to speech perception and subsequent spoken language understanding.

Comprehension and Communicability Tests

A more global assessment of the performance of a communication system is the listener's ability to understand and respond to selected aspects of the linguistic message. Comprehension tests typically involve answering questions or verifying statements about the semantic or pragmatic content of speech produced in short sentences. The primary dependent variable in these tests is response latency (or delay), since accuracy scores are nearly perfect when the speech is presented under clear conditions. Under less favorable conditions, accuracy and response latencies may trade off, depending on the subject's criterion.

Communicability tests typically use a two-way communication task in which pairs of subjects interact using a given transmission system (Schmidt-Nielsen and Everett, 1982; Schmidt-Nielsen, 1985). In some tests the subjects are required to carry out a specific task that involves the active exchange of information to solve a problem interactively. These tasks are often referred to as utility tests because the communication system is being used as it might be used in an actual application. The major advantage of a two-way test is that, because it is interactive, the talkers can adapt to the transmission system by talking more loudly or clearly, for example, if needed. This approach to assessing speech perception may be constrasted with the more traditional one-way tests that are concerned with testing the limits of a system to transmit attributes of the speech signal in isolation using speech intelligibility or speech quality tests.

Tests of Listener Fatigue, Workload, and Processing Capacity

A variety of experimental techniques can be used to measure the amount of processing capacity and resource allocation required for a given task. The results of these tests are typically interpreted as some index of the amount of mental effort or attention required by a listener to carry out a specific task. Many of the procedures involve memory tasks in which subjects are required to memorize and subsequently recall lists of words under different recall conditions (Rabbitt, 1966, 1968; Luce, Feustel, and Pisoni, 1983). Other techniques involve dual tasks in which the human operator is required to carry out at least two simultaneous tasks (Kantowitz and Sorkin, 1983).

A number of experiments have measured the accuracy and speed of performance when additional tasks are added to the primary task. Primary and secondary task performance is assessed as a function of the difficulty of both task components (Wickens, 1984). In addition to measures of cognitive workload based on performance, worker opinion questionnaires, interviews, and rating scales have also been employed. Some researchers have attempted to develop physiological measures of mental workload based on heartbeat rhythm and on components of the cortical evoked potential. At the present time, however, there are no standards for assessing workload that have been generally accepted (Gopher and Donchin, 1986).

Speech Quality Tests

Numerous procedures and tests have been developed over the years to measure speech quality and to quantify some of the more prominent subjective attributes or dimensions of speech (Munson and Karlin, 1962; Hawley, 1977; Nickerson and Huggins, 1977; Voiers, 1977a; Woodward and Cupples, 1983; Kitawaki, Honda, and Itoh, 1984). The most common of these tests involve rating scales, questionnaires, or direct paired-comparisons that are used to elicit magnitude estimations or scaling responses from listeners. The scores obtained from these tests are often subjected to multidimensional scaling analyses to obtain psychological dimensions that can be related to physical properties of the particular systems under study (Shepard, 1972; Wish and Carroll, 1974). Although often very highly correlated with each other, speech intelligibility and speech quality tests are assumed to measure different attributes of speech.

The human observer is an extremely flexible processor of information and is able to adapt his or her behavior quickly to specific task demands so as to produce optimal or near-optimal performance despite variation in certain irrelevant dimensions. In the case of traditional speech intelligibility tests, for example, listeners are often able to ignore stimulus attributes related to quality and naturalness and direct their attention to the acoustic-phonetic properties of the speech signal that distinguish minimal pairs of words. Similarly, listeners are able to ignore many attributes related to phonetic intelligibility and focus their attention instead on judgments of speech quality or preference. The degree to which speech intelligibility and quality attributes can be selectively ignored by the listener depends largely on the specific demands of the listening task.

EVALUATION OF NOISE REDUCTION METHODS

Over the years, a number of noise reduction methods have been developed to improve the quality and the intelligibility of speech degraded by background noise, and a number of perceptual tests have been carried out to assess the effects of those methods on listener performance. Lim and Oppenheim (1979) have reviewed many of these methods and have reported the results of speech discrimination and speech quality tests. The general conclusion that emerged from their survey was that, although many of the proposed noise reduction systems obtained higher ratings in speech quality in terms of reduced perception of noisiness, there was little if any corresponding improvement in speech intelligibility scores for familiar materials. Indeed, of the systems reviewed, almost all of them actually reduced intelligibility scores, and those that did not appeared to introduce other degradations in speech quality, as previously mentioned. Thus, the results of their survey were less than encouraging and led them to conclude that a great deal more basic research remained to be done before any improvements would be observed.

Since the Lim and Oppenheim (1979) review, research has continued on the development of noise reduction methods; several methods were described in the previous major section.

The Rome Air Development Center (RADC) speech enhancement unit (SEU), developed by Weiss and Aschkenasy (1978), was tested in an Air Force operational radio communication environment having a relatively low speech-to-noise ratio (Woodard and Cupples, 1983). The unit was evaluated in several ways using traditional speech intelligibility tests (DRTs), a speech quality evaluation test measuring readability (i.e., overall speech quality) using a five-point subjective rating scale, and copying-time measures. The results of these tests using the SEU showed no overall increase in DRT intelligibility scores, an increase of one point in the five-point speech quality (readability) scores, and a decrease in message copying times. These findings were interpreted as improvements in human processing efficiency and, therefore, as decreases in operator workload with the processing unit in use (Cupples and Foelker, 1987).

Representative of the frequency-selective noise reduction methods is that of McAulay and Malpass (1980), which showed no improvement in intelligibility as measured by the closed-response DRT (Sandy and Parker, 1982).

Some performance tests have been conducted recently with a noise reduction system for hearing aids, the Zeta Noise Blocker (a chip-level hardware version of the Wiener filter described above), developed by Graupe et al. (1987). Speech intelligibility tests using the Northwestern University monosyllabic word lists under five noise conditions were carried out by Stein and Dempsey-Hart (1984). Depending on the noise spectra, statistically reliable increases in intelligibility were observed in both normal-hearing listeners and listeners with a sensorineural hearing loss. The largest improvements in intelligibility scores were observed with low-frequency noise (600 to 800 Hz) and with a noise referred to as "cafeteria noise," the characteriestics of which were not specified. No improvement was observed with noise having a flat spectrum.

Another study of the Zeta Noise Blocker was carried out by Wolinsky (1986) using 18 subjects with moderate to severe sensorineural hearing loss. Monosyllabic words from the Northwestern University lists were presented in four noise conditions. Of the 18 patients, 17 displayed statistically significant improvements using the Zeta adaptive filter in one or more of the noise conditions. As in the previous study, the greatest improvements were observed in the low-frequency noise and the "cafeteria noise" conditions, although some patients showed improvements in the higher-frequency noise condition. In both studies (Stein and Dempsey-Hart, 1984; Wollinsky, 1986), subjects were allowed to adjust the volume controls of their hearing aids in both filter-on and filter-off conditions. It is therefore not possible to determine unequivocally what proportion of the observed improvements were attributable to the action of the Zeta Noise Blocker, the change in overall hearing aid gain that would result from adjustment of the volume control, or the interaction between these two factors.

In addition to these studies, anecdotal evidence from some highly experienced listeners suggests that noise reduction techniques may be helpful in analyzing speech from noisy recordings. Despite the absence of quantitative evidence to support their conclusions, expert listeners consistently report that noise reduction methods and other signal processing techniques appear to be beneficial in performing transcription tasks with poor quality speech (see the appendix). Other experiments have also shown, however, that listeners can learn to listen to severely distorted speech and increase their comprehension of it (Tobias and Irons, 1973).

Finally, no independent evaluations comparing two or more noise reduction methods under similar conditions have been reported in the literature. This situation makes it difficult, if not impossible, to choose among competing methods on an objective basis.

CONCLUSIONS

Performance data on the evaluation of noise reduction devices present a mixed picture. No improvement in speech intelligibility has ever been reported when a test using a closed-response set has been used to evaluate speech intelligibility. The most frequently used test is the diagnostic rhyme test, which involves discriminating between only two response alternatives. Balanced against these negative test results are the findings obtained in evaluations based on purely subjective judgments of speech quality. These assessments almost always show that noise reduction devices produce improved ratings of speech quality. Trying to resolve this conflict became a major preoccupation of the panel. In recent intelligibility tests, using an open-response format, increases in intelligibility have been observed for at least one noise reduction method under a few noise conditions for both normal and pathological ears. However, several reservations were noted about the experimental procedures used in these studies.

The panel concludes that any increases in intelligibility that may result from the use of existing noise reduction techniques are sufficiently small in magnitude that they cannot be measured by closed-response tests such as the DRT. The panel therefore recommends the use and development of intelligibility testing methods with higher sensitivity, such as open-response tests, to assess speech intelligibility with noise reduction methods. Promising noise reduction techniques should then be tested using these intelligibility tests. Since no studies have been reported to assess the effects of noise reduction techniques on listener fatigue, workload, or mental effort, such studies obviously need to be performed as well. In general, the panel believes that noise reduction devices may be beneficial in many situations and recommends that more extensive assessment be instituted.

Conclusions and Recommendations

After reviewing and evaluating the effectiveness of several techniques designed to remove noise from noise-degraded speech signals, the panel reached a number of conclusions.

CONCLUSIONS

(1) Some noise reduction methods at present appear to be useful in improving the quality of speech in noise.

(2) No improvement in human speech intelligibility has been demonstrated by closed-response tests such as the diagnostic rhyme test.

(3) Other methods for testing speech intelligibility may be more appropriate for assessing noise reduction techniques.

(4) The literature examined by the panel showed that no formal studies have been reported that assess the effects of noise reduction methods on listener fatigue, workload, or mental effort.

(5) Different noise reduction techniques have been designed to be most effective for different types of noise.

(6) Reduction of continuous noise is obtained chiefly as a result of adjustment of the short-term (20 to 40 ms) spectral magnitude of the noisy speech signal, rather than by adjustment of the short-term phase.

(7) Mathematical and statistical criteria have mainly been used in the design of noise reduction methods with minimal regard for perceptual criteria derived from studies using human listeners.

(8) Noise reduction is enhanced by the use of short-term time-dependent decisions, such as estimation of the noise spectrum, temporal speech characteristics, and type of speech (e.g., voiced or unvoiced).

The above conclusions lead to several corollary recommendations, which point to the need for a program of basic and applied research to develop new testing procedures that are more appropriate for evaluating noise reduction methods and to reassess promising noise reduction methods using the new testing procedures. For developing improved noise reduction techniques, the panel stresses the need for discovery and mathematical formulation of properties of human speech perception in noise, to be used in the derivation of new perceptually based design criteria. Specifically, the Panel makes six recommendations.

RECOMMENDATIONS

(1) The use of new testing procedures, especially open-response tests, for evaluating noise reduction methods for a variety of noise types should be explored.

(2) These tests should not only assess speech intelligibility, but should also assess speech quality, fatigue, workload, and mental effort.

(3) Additional basic research should be carried out to specify the interrelationships among measures of speech intelligibility, speech quality, fatigue, etc.

(4) Promising noise reduction techniques should be reassessed once the above studies have been carried out.

(5) In developing improved noise reduction techniques, judicious use should be made of additional known short-term properties of speech signals, noise, and human auditory perception.

(6) Research should be performed to further our knowledge of human speech perception in noise. On the basis of a mathematical formulation of this knowledge, perceptually based criteria should be derived for the design of new noise reduction methods.

Appendix
Experiences of an Expert Listener

One of the members of the panel, Thomas Stockham, has had many years' experience as a part-time tape analyst and transcriber. This experience, which began with his participation as one of the six members of the advisory panel on White House tapes appointed by Chief Judge John J. Sirica in 1973, has involved a wide variety of real-life tapes. These tapes have brought him in contact with a number of speech recordings in which the speech was embedded in quantities of noise that seriously compromised the intelligibility or of the transcribability of the speech signal. Over the last 14 years, Stockham transcribed approximately 15 tapes averaging five or so minutes in duration. As is usually the case, there was never a ground-truth transcript to compare with the results. Nevertheless, certain aspects of the transcribing experience, the way the meanings of the transcript evolve, and occasional and informal interactions with the conversational participants convinced him that the technical aids used for enhancing the accuracy and completeness of such transcripts are distinctly effective. Stockham estimates that an additional 5 to 20 percent of the text can be transcribed, depending on the methods used and the signal being transcribed, with the increase often containing pivotal material. By virtue of his working habits, Stockham has had occasion to notice an important phenomenon. After much listening time without technical aids had produced a transcript, the introduction of technical aids almost always produced a new flurry of discovery. Although possible, it seems unlikely to him that this new discovery comes from heightened perception induced by the anticipation of what the new method might bring. Neither does it seem likely to him that the new discovery is a chance phenomenon.

While he always works alone, Stockham has often had occasion to share his experiences concerning each transcription with others. There is general agreement concerning the independently discovered advantages of the methods described below. In addition, there seems to be a protocol for listening that is common to the habits of listeners who work in this area. Indeed, in these discussions the ranking of the effectiveness of many techniques among different workers seems to be similar.

In the course of trying to develop as complete a transcript as possible while seeking high accuracy, Stockham employed a collection of signal processing and playback methods as aids to understanding the speech. These methods spanned the spectrum from highly sophisticated computer-based algorithms to the relatively simple use of tone controls, tape speedup or slowdown, etc. These analyses generally place a high premium on accuracy with essentially no time constraints. It seems likely that these experiences apply broadly, and

Stockham is convinced that the use of these methods materially aids in the production of an accurate transcript.

The methods Stockham used in producing as complete a transcript as possible from a noisy tape recording involved not only the processing of the recorded signal by various methods, but also a discipline of listening and listening practice. The processing methods involved a variety of computer-based techniques for the removal of noise, reverberation, resonance, tones, and static. (Some of these techniques are similar to those examined in this report.) In addition to these, he used real-time hardware such as highpass and lowpass tunable filters, octave and one-third-octave-band equalizers, accurate tone controls, volume compressors and expanders, variable-speed tape players, computer-based random access digital audio editors, and sound-conditioned listening rooms. All of these were used in a variety of combinations or separately. As part of the listening discipline, a variety of practices have been found useful in increasing productivity:

- Listening at high levels and low levels.
- Repetitive listening to single sentences or important phrases.
- Enforced periods of rest.
- Alternation of concentrated examination of critical sections with global review of the entire recording.

In addition, it is clear to Stockham that regular listening practice was an important ingredient in the productivity and accuracy of his efforts. This practice was important regardless of the number of technical aids employed. In conclusion, Stockham's judgment as an experienced listener is that software and hardware processing of noise-contaminated speech recordings can almost certainly produce transcripts superior to those that can be obtained without their use.

References

American National Standards Institute

1969 Methods for the Calculation of the Articulation Index. ANSI S3.5-1969.

Aschkenasy, E.

1986 Multichannel Advanced Speech Enhancement Development. Technical Report RADC-TR-86-244, Rome Air Development Center, Griffiss AFB, N.Y.

Baddeley, A.

1986 *Working Memory*. Oxford, UK: Oxford University Press.

Beek, B., Neuburg, E., and Hodge, D.

1977 An assessment of the technology of automatic speech recognition for military applications. *IEEE Trans. Acoustics, Speech, and Signal Processing* ASSP-25(4):310-320.

Berouti, M., Schwartz, R., and Makhoul, J.

1979 Enhancement of speech corrupted by acoustic noise. *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing* 208-211.

Boll, S.F.

1979 Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoustics, Speech, and Signal Processing* ASSP-27(2):113-120.

Chabries, D.M., Christiansen, R.W., Brey, R.H., and Robinette, M.S.

1982 Application of the LMS adaptive filter to improve speech communication in the presence of noise. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* 148-151.

Chabries, D.M., Christiansen, R.H., Brey, R.H., Robinette, M.S., and Harris, R.W.

1987 Application of adaptive digital signal processing to speech enhancement for the hearing impaired. *Journal of Rehabilitation Research and Development* 24(4):65-74.

Cupples, E.J., and Foelker, J.L.

1987 Air Force speech enhancement program. *Proceedings of Military Speech Tech-87*. New York: Media Dimensions.

Ephraim, Y., and Malah, D.

1984 Speech enhancement using a minimum-mean-squared-error short-time spectral amplitude estimator. *IEEE Trans. Acoustics, Speech and Signal Processing* ASSP-32(2):1109-1121.

French, N.R., and Steinberg, J.C.

1949 Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America* 19:90-119.

Gopher, D., and Donchin, E.

1986 Workload - An examination of the concept. Chapter 41 in K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance*, Vol.II. New York: John Wiley.

Graupe, D., ed.

1984 *Time Series Analysis, Identification and Adaptive Filtering*. Chapters 10 and 11. Malabar, Florida: Krieger Publishing Company.

Graupe, D., Grosspietsch, J.K., and Basseas, S.P.

1985 A self adaptive filter of environmental noises from speech and its evaluation. Pp. 1261-1263 in *Proc. 24th Conf. on Decision and Control*.

1987 A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation. *Journal of Rehabilitiation Research and Development* 24(4).

Hawley, M., ed.
1977 *Benchmark Papers in Acoustics,* Volume 11: *Speech Intelligibility and Speaker Recognition.* Stroudsburg, Pennsylvania: Dowden, Hutchinson and Ross.

Hirsh, I.J., Davis, H., Silverman, S.R., Reynolds, E.G., Eldert, E., and Benson, R.W.
1952 Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders* 17:321-337.

Hockey, R., ed.
1983 *Stress and Fatigue in Human Performance.* Chichester, UK: Wiley.

House, A.S., Williams, C.E., Hecker, M.H., and Kryter, K.D.
1965 Articulation testing methods: Consonantal differentiation with a closed response set. *Journal of the Acoustical Society of America* 37:158-166.

Huggins, A.W., and Nickerson, R.S.
1985 Speech quality evaluation using phoneme-specific sentences. *Journal of the Acoustical Society of America* 77:1896-1906.

Kalikow, D.N., Stevens, K.N., and Elliott, L.L.
1977 Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61:1337-1351.

Kantowitz, B.H., and Sorkin, R.D.
1983 *Human Factors: Understanding People-System Relationships.* New York: John Wiley.

Kidd, G., Mason, C.R., and Green, D.M.
1986 Auditory profile analysis of irregular sound spectra. *Journal of the Acoustical Society of America* 79:1045-1053.

Kitawaki, N., Honda, M., and Itoh, K.
1984 Speech-quality assessment methods for speech-coding systems. *IEEE Communications Magazine* 22(10):26-33.

Kryter, K.D.
1962 Methods for the calculation and use of the articulation index. *Journal of the Acoustical Society of America* 34:1689-1697.
1985 *The Effects of Noise on Man.* Orlando, Florida: Academic Press.

Kundu, A., and Mitra, S.K.
1987 A computationally efficient approach to the removal of impulse noise from digitized speech. *IEEE Trans. Acoustics, Speech and Signal Processing* ASSP-35(4):571-574.

Lim, J.S.
1978 Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise. *IEEE Trans. Acoustics, Speech, and Signal Processing* ASSP-26(5):471-472.

Lim, J.S., ed.
1983 *Speech Enhancement.* Englewood Cliffs, New Jersey: Prentice-Hall.

Lim, J.S., and Oppenheim, A.V.
1979 Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE* 67(12):1586-1604.

Loeb, M., ed.
1986 *Noise and Human Efficiency.* Chichester, UK: Wiley.

Luce, P.A., Feustel, T.C., and Pisoni, D.B.
1983 Capacity demands in short-term memory for synthetic and natural word lists. *Human Factors* 25:17-32.

McAulay, R.J., and Malpass, M.L.
1980 Speech enhancement using a soft-decision maximum likelihood noise suppression filter. *IEEE Trans. Acoustics, Speech, and Signal Processing* ASSP-28(2):137-145.

Miller, G.A., Heise, G.A., and Lichten, W.
1951 The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology* 41:329-335.

Miller, G.A., and Nicely, P.E.
1955 An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27:338-352.

Munson, W.A., and Karlin, J.E.
1962 Isopreference method for evaluating speech-transmission circuits. *Journal of the Acoustical Society of America* 34:762-774.

Nakatsui, M.
1979 Subjective evaluation of SPAC in improving quality of noisy speech. *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing* 467-470.

Nickerson, R.S., and Freeman, B.
1974 Discrimination of the order of the components of repeating tone sequences: Effect of frequency separation and extensive practice. *Perception and Psychophysics* 3:471-477.

Nickerson, R.S., and Huggins, A.W.F.
1977 The Assessment of Speech Quality. BBN Technical Report No. 3486. Bolt Beranek and Newman Inc., Cambridge, Massachusetts.

Oppenheim, A.V., and Schafer, R.W., eds.
1975 *Digital Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall.

Paul, J.E.
1978 Automatic digital audio processor (ADAP). *Proc. 11th Annual Asilomar Conf. on Circuits, Systems, and Computers*. Pacific Grove, California.

Picheny, M.A., Durlach, N.I., and Braida, L.D.
1985 Speaking clearly for the hard-of-hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research* 28:96-103.

Pisoni, D.B., Nusbaum, H.C., and Greene, B.G.
1985 Perception of synthetic speech generated by rule. *Proc. of the IEEE* 73:1665-1676.

Preuss, R.D.
1979 A frequency domain noise-cancelling preprocessor for narrowband speech communications systems. *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing* 212-215.

Rabbitt, P.
1966 Recognition memory for words correctly heard in noise. *Psychonomic Science* 6:383-384.
1968 Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology* 20:241-248.

Sambur, M.R.
1978 Adaptive noise cancelling for speech signals. *IEEE Trans. Acoustics, Speech and Signal Processing* ASSP-26(5):419-423.

Sandy, F., and Parker, J.F.
1982 Digital Voice Processor Consortium Interim Report. Report No. MTR-81w00159-01, The MITRE Corp., Bedford, Massachusetts.

Schmidt-Nielsen, A.
1985 Problems in evaluating the real-world suitability of digital voice communication systems. *Behavior Research Methods, Instruments, and Computers* 17(2):226-234.
1987 Evaluating degraded speech: Intelligibility tests are not all alike. *Proceedings of Military Speech Tech-87*. New York: Media Dimensions.

Schmidt-Nielsen, A., and Everett, S.S.
1982 A conversational test for comparing voice systems using two-way communication links. *IEEE Trans. Acoustics, Speech, and Signal Processing* ASSP-30:853-863.

Schroeder, M.R., and Noll, A.M.
1965 Recent studies in speech research at Bell Telephone Laboratories. Paper A21. *5th International Congress of Acoustics*, Liege, Belgium.

Shepard, R.N.
1972 Psychological representation of speech sounds. Pp. 67-113 in E.E. David and P.B. Denes, eds., *Human Communcation: A Unified View*. New York: McGraw-Hill.

Shiffrin, R.M.
1976 Capacity limitations in information processing, attention, and memory. In W.K. Estes, ed., *Handbook of Learning and Cognitive Processes*, Vol. 4. Hillsdale, New Jersey: Erlbaum.

Steele, R., and Goodman, D.J.
1977 Detection and selective smoothing of transmission errors in linear PMC. *Bell System Technical Journal* 56:369-409.

Steeneken, H.J.M., and Houtgast, T.
1980 A physical method for measuring speech transmission quality. *Journal of the Acoustical Society of America* 67:318-326.

Stein, L.K., and Dempsey-Hart, D.
1984 Listener-assessed intelligibility of a hearing aid self-adaptive noise filter. *Ear and Hearing* 5:199-204.

Stevens, K.N.
1964 Acoustical aspects of speech production. Pp. 347-355 in W.O. Fenn and H. Rahn, eds., *Handbook of Physiology-Respiration I*. American Physiology Society.

Suzuki, J.
1976 Speech processing by splicing of autocorrelation function. *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing* 113-116.

Suzuki, J., Igarashi, J., and Ishii, Y.
1977 Extraction of speech in noise by digital filtering. *Journal of the Acoustical Society of Japan* 33(8):405-411.

Tillman, T.W., and Carhart, R.
1966 Northwestern University Auditory Test No. 6. Report No. USAF-SAM-TR-66-55. School of Aerospace Medicine, United States Air Force. Published in H. Davis and S.R. Silverman, eds. (1970), *Hearing and Deafness*. New York: Holt, Rinehart, and Winston.

Tobias, J.V., and Irons, F.M.
1973 Reception of Distorted Speech. Report No. FAA-AM-73-13. Federal Aviation Administration, Office of Aviation Medicine, Civil Aeromedical Institute, Oklahoma City, Oklahoma.

Van Trees, H.L.
1968 *Detection, Estimation and Modulation Theory,* Part I. New York: Wiley.

Veenemand, D.E., and Mazor, B.
1987 Enhancement of block-coded speech. *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing* 193-196.

Voiers, W.D.
1977a Diagnostic acceptability measure for speech communication systems. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* 204-207.
1977b Diagnostic evaluation of speech intelligibility. Pp. 374-387 in M. Hawley, ed., *Benchmark Papers in Acoustics,* Vol. 11: *Speech Intelligibility and Speaker Recognition.* Stroudsburg, Pennsylvania: Dowden, Hutchinson and Ross.
1983 Evaluating processed speech using the Diagnostic Rhyme Test. *Speech Technology* (Jan.-Feb.):30-39.

Wang, D.L., and Lim, J.S.
1982 The unimportance of phase in speech enhancement. *IEEE Trans. Acoustics, Speech, and Signal Processing* ASSP-30(4):679-681.

Wang, M.D., and Bilger, R.C.
1973 Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America* 54:1248-1266.

Watson, C.S., Kelly, W.J., and Wroton, H.W.
1976 Factors in the discrimination of tonal patterns II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America* 60:1176-1186.

Webster, J.C.
1978 Speech interference aspects of noise. Pp. 193-228 in D.M. Lipscomb, ed., *Noise and Audiology.* Baltimore, Maryland: University Park Press.

Weiss, M., and Aschkenasy, E.
1975 Study and Development of the INTEL Technique for Improving Speech Intelligibility. Technical Report RADC-TR-75-108. Rome Air Development Center, Griffiss AFB, New York.
1978 The Speech Enhancement Advanced Development Model. Technical report RADC-TR-78-232. Rome Air Development Center, Griffiss AFB, N.Y.

Wickens, C.D.
1984 *Engineering Psychology and Human Performance.* Columbus, Ohio: Charles E. Merrill.

Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Hearn, R.H., Zeidler, J.R., Dong, E., and Goodlin, R.C.
1975 Adaptive noise cancelling: Principles and applications. *Proc. of the IEEE* 63(12):1692-1716.

Wish, M., and Carroll, D.
1974 Applications of individual differences scaling to studies of human perception and judgment. Pp. 449-491 in E.C. Carterette and M.P. Friedman, eds., *Handbook of Perception.* New York: Academic Press.

Wollinsky, S.
1986 Clinical assessment of a self-adaptive noise filtering system. *The Hearing Journal* (Oct.):29-32.

Woodard, J.P., and Cupples, E.J.
1983 Selected military applications of automatic speech recognition technology. *IEEE Communications Magazine* 21(9):35-41.

Technical Reports Distribution List

Dr. Curt Boylls
National Security Agency
Attn: R523
9800 Savage Road
Ft. George C. Meade MD 20755

Dr. Allen Boysen
Audiology & Speech Pathology Service
Dept. of Veterans Affairs (126CO)
810 Vermont Avenue, N. W.
Washington, DC 20420

Dr. James E. Densmore, Director
Office of Environment and Energy
AEE-1
Federal Aviation Administration
800 Independence Avenue, S. W.
Washington, DC 20591

Capt. Thomas N. Jones, MSC, USN
Naval Medical Research and Development Command
National Medical Center
Bethesda, MD 20814

LtC. Bruce Leibrecht, USA
U. S. Army Aeromedical Research Laboratory
Attn: SGRD-UAS/LtC. Leibrecht
P. O. Box 577
Ft. Rucker, AL 35330

Dr. Richard Loutitt, Director
Division of Behavioral & Neural Science Room 320
National Science Foundation
1800 G Street, N. W.
Washington, DC 20550

Dr. Andrew Monjan
Neuroscience and Neuropsychology of Aging Program
National Institute of Aging
Building 31, Room 5C-35
Bethesda, MD 20205

Ralph F. Nauton, MD, FACS
Acting Director, Extramural Programs
NIDCD
Federal Building, Room 1C-11
Bethesda, MD 20892

Dr. Arnauld Nicogossian
Office of Occupational Medicine
NASA Headquarters, Code NIH-34
Washington, DC 20546

Dr. Charles W. Nixon
USAF AAMRL - BBA
Wright-Patterson AFB, OH 45433

Dr. John J. O'Hare
Office of Naval Research
Code 1142PS Rm 823
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Christopher Platt
Sensory Systems, BNS, Room 320
National Science Foundation
1800 G Street, N. W.
Washington, DC 20550

Dr. Clemans A. Powell
NASA Langley Research Center
Mail Stop 463
Hampton, VA 23665

Commander
U.S. Army Medical Research and Development Command
Attn: SGRD-PCL/Col. Sedge
Ft. Detrick
Frederick, MD 21701

Dr. Kevin Shepherd
NASA Langley Research Center
Mail Stop 463
Hampton, VA 23665

Dr. Janis Stoklosa
NASA Headquarters/EBM
Washington, DC 20546

Dr. John F. Tangney
AFOSR/NL
Building 410
Bolling AFP DC 20332

Mr. Richard M. Tedrick, Chief
International and Scientific Program Office, AEE-3
Federal Aviation Administration
800 Independence Avenue, S. W.
Washington, DC 20591

Dr. Thomas Tredici, MC, USAF
USAF School of Aerospace Medicine
Brooks AFB, TX 78235

Dr. Milton A. Whitcomb
NAS-National Research Council
CHABA
2101 Constitution Avenue, N.W.
Washington, DC 20418

Dr. Carl E. Williams
Naval Aerospace Medical Research Laboratory
Naval Air Station
Pensacola, FL 32512

Col. John Wolcott, USAF
Deputy Commander for Research, Development, & Acquisition
Aerospace Medical Division
Brooks AFB, TX 78235

Dr. James W. Wolfe
NASA Suite 600
600 Maryland Avenue, S. W.
Washington, DC 20024

Capt. James N. Woody, MC, USN
Commanding Officer
Naval Medical Research and Development Command
NMCNCR
Bethesda, MD 20814-5044